



# Algorithm Modified K-Nearest Neighbor (M-KNN) for Classification of Attention Deficit Hyperactive Disorder (ADHD) in Children

Masdiana Sagala

Faculty of Computer Science, Universitas Katolik Santo Thomas, Indonesia

## Article Info

### Article history:

Received, Apr 03, 2019

Revised, Apr 29, 2019

Accepted, May 27, 2019

### Keywords:

ADHD,  
score,  
classification,  
k-nearest neighbor,  
modified k-nearest neighbor.

## ABSTRACT

Childhood is a time when they observe everything around them to learn, experience and grow. Children are human resources that must be prepared early on to be able to develop optimally according to their abilities, but not every child is born in normal conditions. The MKNN method was implemented to determine the potential for ADHD by using 3 data parameters, namely age, gender, and Score Score. In this study, the Modified k - Nearest Neighbor (MKNN) classification technique was used by using two criteria, potentially ADHD and not potentially ADHD. MKNN is a modification of the k classification technique - Nearest Neighbor (KNN) which will be used to process the features of existing data into a decision. The modification of this method aims to overcome the weakness regarding the distance of the data with the weight which has many problems in outliers in the KNN method. This research was conducted using 10 total datasets. Each number of datasets will be randomly divided into training data and test data. Testing of this system is carried out with a k value of 3. And the average accuracy is obtained. The average maximum accuracy value generated by the system is 50% when the number of datasets is 10.

*This is an open access article under the [CC BY-SA](#) license.*



## Corresponding Author:

Masdiana Sagala,  
Faculty of Computer Science, Universitas Katolik Santo Thomas,  
Universitas Katolik Santo Thomas,  
Jl. Setia Budi, Kp. Tengah, Kec. Medan Tuntungan, Kota Medan, Sumatera Utara 20135.  
Email: masdianasagala@gmail.com

## 1. INTRODUCTION

Concentration and hyperactivity disorder, often referred to as Attention Deficit Hyperactive Disorder (ADHD), is a neuropsychiatric syndrome recently found in children, usually accompanied by symptoms of hyperactivity and impulsive behavior[1]. The main disadvantage that a child with ADHD experiences is a striking barrier between themselves and the consequences that accompany their life. This highlights the problem of children with ADHD who are always considered uncooperative and very naughty[2]. ADHD children do not respond when given directions in the same way as other children, due to their lack of ability to concentrate and respond to tasks or activities.

In this study, the authors tried to classify attention deficit disorder and hyperactivity which is often referred to as Attention Deficit Hyperactive Disorder (ADHD) which often occurs in children from 5-14 years of age, 4 times more sufferers in boys. than girls. Three types of ADHD are classified as common or many occur in children in general. However, the author will try to dig up more information about various types of diseases that fall into the scope of ADHD[3].

Classification is a technique of mapping (classifying) data into one or more predefined classes[4]. There are many classification techniques that can be used, including Naïve Bayes, k-Nearest Neighbor, and Artificial Neural Network. The use of the right algorithm can increase the accuracy of the decisions taken. The classification method used by the author is the Modified k-Nearest Neighbor (MKNN)[5].

With the K-Nearest Neighbor (KNN) method, the Modified K-Nearest Neighbor (MKNN) method was developed. The main idea of this method is to classify test samples according to neighbor tags. This method is a kind of weighted KNN so that the weights are determined using a different procedure. The procedure for calculating the fraction of labeled neighbors is equal to the total number of neighbors. Because the level of accuracy of the Modified K-Nearest Neighbor (MKNN) method is better when compared to the K-Nearest Neighbor (KNN) method. Like previous experiments on the Wine dataset, the KNN method has an accuracy rate of 83.79% while the MKNN method is 85.76% and also on the Isodata dataset the method has an accuracy rate of K-Nearest Neighbor (KNN) 82.90% while the Modified K-Nearest Neighbor method (MKNN) 83.32% [6].

## 2. RESEARCH METHOD

In carrying out this research, clear and structured stages are needed, in order to facilitate the process, it is necessary to make a diagram design such as the diagram below:

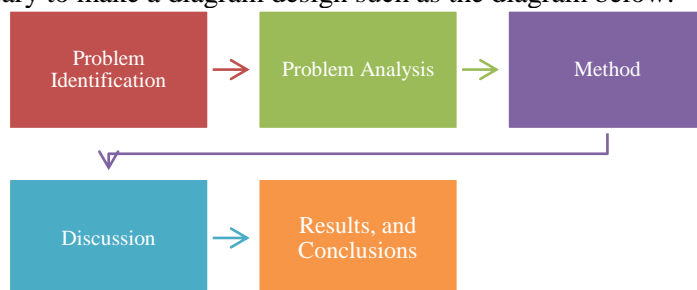


Figure 1. Diagram of Methods and Research Stages

In the stages of the research method, the author conducted interviews with experts to obtain symptoms of worms in livestock.

### 2.1. Basic theory

#### A. Artificial intelligence

Artificial intelligence comes from the English word "Artificial Intelligence" or AI for short, namely intelligence is an adjective which means intelligent, while artificial means artificial, the intelligence referred to here refers to machines that are able to think, weigh actions to be taken, and are able to make decisions such as done by humans. Such a system is generally considered a computer. Intelligence is created and fed into a computer machine so that it can do work as humans can[7].

Artificial intelligence is the idea of making computer software that has intelligence so that computer software can do a job done by humans. The job is in the form of consultation which can provide information in the form of suggestions that will be very useful. Artificial intelligence allows computers to think in a way that simplifies programs. In this way, artificial intelligence can mimic the human learning process so that new information can be absorbed and used as a reference in the future[8].

#### B. K-Nearest Neighbor (KNN)

The k-nearest neighbor (KNN) algorithm is a method for classifying objects based on learning data that is closest to the object. This technique is very simple and easy to implement. Similar to the clustering technique, grouping new data based on the distance of the new data to some of the data / closest neighbors. To define the distance between two points, namely the point on the training data (x) and the point on the testing data (y), the Euclidean formula is used, as shown in the equation below[9] :

$$d_{(x_1, x_2)} = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \dots \dots \dots (1)$$

where  $x_1$  and  $x_2$  are two data with  $n$  attributes. Normalization is done to prevent attributes that have too large a range with attributes that are of small value. The normalization min-max calculation can be used to change the attribute  $A$  with a value of  $v$  to a value of  $d'$  in the range  $[0,1]$ . Calculation of the min-max normalization with the equation:

$$d' = \frac{(d - \min_{euc})}{(\max_{euc} - \min_{euc})} \dots \dots \dots (2)$$

Where:

$d'$  = normalized value

$d$  = data value to be normalized

$\min_{euc}$  = the lowest (minimum) value of data on attribute  $A$

$\max_{euc}$  = the highest (maximal) value of data on attribute  $A$ .

The main idea of this method is to enter class labels from the data based on the data points on the validated training data with  $k$  values. In other words, first, the data validity was calculated on all the data in the training data. Furthermore, the calculation is carried out to find the Weight Voting on all test data using data validity[10].

The following is the Modified KNN algorithm:

```

Output_label := MKNN (train_set, test_sample )
Begin
For i := 1 to train_size
Validity(i) := Compute Validity of i-th sample;
End for;
Output_label := Weighted_KNN(Validity, test_sample);
Return Output_label ;
End.

```

### C. Attention Deficit Hyperactive Disorder (ADHD)

The American Psychiatric Association (APA) in the DSM-IV-TR categorizes ADHD into three types, namely[3]:

1. ADHD with attention characteristics, namely:
  - a. Often unable to pay attention to small things or make mistakes, not being thorough in schoolwork, work, or other activities.
  - b. Often have difficulty maintaining attention in carrying out tasks or play activities.
  - c. Often seems inconsiderate when speaking directly.
  - d. Often does not follow instructions and fails to complete school assignments, daily tasks, or duties at work
  - e. Often has difficulty organizing tasks and activities.
  - f. Often refuses, dislikes, or is reluctant to take part in tasks that require constant mental effort (e.g. schoolwork or homework)
  - g. Often loses things needed to do assignments or other activities (for example toys, schoolwork, pencils, books, or other tools).
  - h. Often easily distracted by foreign stimuli.
  - i. Often times forgetting in daily activities.
2. ADHD with hyperactive characteristics, namely:
  - a. Often fidgeting with hands or feet or writhing in a chair.
  - b. Often leaving a seat in class or in other situations requires sitting quietly.
  - c. Often running around or climbing excessively in situations that deem it inappropriate.
  - d. Often has difficulty playing or participating in pleasant activities in a quiet manner.
  - e. Often in a hurry or in constant motion as if pushed by a machine.
  - f. Often talk too much.
3. ADHD whose impulsivity characteristics are:
  - a. Often answer questions without thinking before the question is finished.

- b. Often has difficulty waiting for their turn.
- c. Frequently interrupting or coercing others (interrupting a conversation and insisting on playing)

### 3. RESULTS AND DISCUSSION

In the description of the data this time, the training data used is the survey data set taken from the results of the questionnaire conducted by the author with 10 respondents as a sample of data. With the provisions of the weight of the assessment, namely:

Table 1. Sample data

SAMPEL	KETERANGAN	BOBOT
SS	SAMA SEKALI TIDAK	0,25
KK	KADANG-KADANG	0,5
SR	SERING	0.75
SL	SELALU	1

Information : Total Value > 21 is identified as Hyperactive, The rest is identified as impulsive.

After the questionnaire was collected, the data would be calculated and the total score for each data sample of each child was calculated, then categorized manually by the researcher, and then calculated using the M-KNN algorithm. The following is a dataset that has been collected by researchers.

Table 2. Questionnaire Dataset

No	Nama	Jenis Kelamin	Usia	Total Nilai	ADHD
1	Naela Cahya	P	11	20,25	Impulsif
2	Dewa Hikmawan	L	5	25	Hiperaktif
3	Pahriansyah	L	6	27,25	Hiperaktif
4	Afrizal Fahrezi	L	12	23	Hiperaktif
5	Muhammad Fayadh	L	7	10,5	Impulsif
6	Nazha Syahira	P	6	17	Impulsif
7	Fithya Zahirah	P	4	12	Impulsif
8	Safitri	P	8	14,75	Impulsif
9	Kaila putri	P	9	22,35	Hiperaktif
10	Khairunnisa	P	11	23	Hiperaktif

Table 3. Manual Calculation Testing Data

1	Naela Cahya	P	11	20,25	Impulsif
2	Dewa Hikmawan	L	5	25	Hiperaktif
8	Safitri	P	8	14,75	Impulsif
9	Kaila putri	P	9	22,35	Hiperaktif

Table 4. Manual Calculation Training Data

3	Pahriansyah	L	6	27,25	Hiperaktif
4	Afrizal Fahrezi	L	12	23	Hiperaktif
5	Muhammad Fayadh	L	7	10,5	Impulsif
6	Nazha Syahira	P	6	17	Impulsif
7	Fithya Zahirah	P	4	12	Impulsif
10	Khairunnisa	P	11	23	Hiperaktif

#### Application of the Modified K-Nearest Neighbor Algorithm

In its application, there are several stages of the Modified K-Nearest Neighbor (MKNN) algorithm, including:

1. Calculating the distance (KNN)

2. Calculating validity
3. Calculating the weighted voting
4. Determine the class category and data accuracy.

#### A. Calculating the distance (KNN)

Data testing 1:

variabel interval scaled				
1	P	11	20,25	Tidak

The calculation of the euclidean distance is used for the age and the total value, as in the example of calculating the first test data according to the euclidean equation which is shown as follows.

$$\begin{aligned}
 d_{(1,3)} &= \sqrt{\sum_{i=1}^n (x_{3i} - x_{1i})^2} \\
 &= \sqrt{(11 - 6)^2 + (20,25 - 27,25)^2} \\
 &= \sqrt{74} \\
 &= 8,6023253
 \end{aligned}$$

Next, calculate the same euclidean distance for all scaled interval variables, and the results are shown in Table 5.

Table 5. Euclidean Data Testing Distance Calculation 1

	SUM Euclidean	Euclidean
d (1,3)	74	8,6023253
d (1,4)	56,5625	7,5208045
d (1,5)	111,0625	10,538619
d (1,6)	46,5625	6,823672
d (1,7)	117,0625	10,819543
d (1,10)	7,5625	2,75

After the euclidean value is obtained, a normalization process is carried out to obtain a distance with a range [0,1] according to the equation shown below.

$$\begin{aligned}
 d'(1,3) &= \frac{(d_{(1,3)} - \min_{euc})}{(\max_{euc} - \min_{euc})} \\
 &= \frac{(8,6023253 - 2,75)}{(10,819543 - 2,75)} \\
 &= 0,7252
 \end{aligned}$$

Next, perform the same normalization calculations for the euclidean distance for all training data, and the results are shown in table 6 below.

Table 6. Results of Normalization Distance Euclidean Data Testing 1

	Euclidean	Normalisasi
d (1,3)	8,6023253	0,7252
d (1,4)	7,5208045	0,5912
d (1,5)	10,538619	0,9651
d (1,6)	6,823672	0,5048
d (1,7)	10,819543	1
d (1,10)	2,75	0

Calculations for categorical variable parameters use the equation with  $p = 1$  and  $d(i, j) = 0$  if the same and 1 if different. The calculation of categorical variables is shown in table 7 below.

Table 7. Result of Calculation of Variable Distance for Data Testing Category 1

d (1,3)	0
d (1,4)	0
d (1,5)	0
d (1,6)	1
d (1,7)	1
d (1,10)	0

After the calculation of the distance for each variable has been generated, the calculation of the average distance will be used by adding the two calculation results and dividing them according to the number of variable types as follows.

$$\begin{aligned}
 d_{(1,3)} &= \frac{d_{var.kategori(1,3)} + d'e(1,3)}{2} \\
 &= \frac{0 + 0,7252}{2} \\
 &= 0,3626
 \end{aligned}$$

Next, calculate the same average distance for all training data, and the results are shown in table 8 below.

Table 8. Distance Calculation Results

	Variabel kategori	Euclidean	Total Jarak
d (1,3)	0	0,7252	0,3626
d (1,4)	0	0,5912	0,2956
d (1,5)	0	0,9651	0,4825
d (1,6)	0	0,5048	0,2524
d (1,7)	1	1	0,5000
d (1,10)	1	0	0

### Calculating Validity

Calculating the validity value of training data with equations. An example of calculating the validity of training data for the first training data ( $x = 1$ ) is as follows.

$$\begin{aligned}
 Validitas (x = 1) &= \frac{1}{k} \sum_{i=1}^k s(label(x), (label(N-i(x)))) \\
 &= \frac{1}{3} \sum_{i=1}^k s(label(x = 1), (label(Ni(x = 2)))) \\
 &= \frac{1}{3} (1 + 0 + 0) \\
 &= \frac{1}{3} \\
 &= 0,333333
 \end{aligned}$$

The same calculation is performed for all training data. The results of this validity calculation are as shown in table 9.

Table 9. Manual Calculation of Data Training Validity

N	k =1	k=2	k=3	Sum S(a,b)	Validitas
3	1	0	0	1	0,333333
4	0	0	0	0	0
5	1	1	0	2	0,666667
6	1	0	1	2	0,666667
7	0	1	1	2	0,666667
1	0	0	0	0	0

After the validity calculation results are complete, the next step is to calculate the weighted voting according to the equation. For example, the calculation of weight voting for data testing 1 is as follows.

$$\begin{aligned}
 w_{1,3} &= Validitas(1) \times \frac{1}{d_{(1,3)} + 0,5} \\
 &= 0,333333 \times \frac{1}{0,3626 + 0,5} \\
 &= 0,3864
 \end{aligned}$$

Next, calculate the same weighted voting for all training data. After obtaining the value of the weighted voting from all training data, a search for the largest weighted voting value is carried out in accordance with the predetermined k value, namely k = 3 and the results are shown in Table 10 below.

Table 10. Calculation Results of Weight Voting Data Testing 1

	Weight Voting	Kategori
d (1,6)	0,8859	Impulsif
d (1,5)	0,8375	Impulsif
d (1,7)	0,6666	Hiperaktif
d (1,3)	0,3864	Impulsif
d (1,4)	0	Hiperaktif
d (1,10)	0	Hiperaktif

In table 10, there are 3 values of weighted voting, namely d (1.6) of 0.8859 with the Impulsive class, d (1.5) of 0.8375 with the Impulsive class, and d (1.7) of 0.666 with Hyperactive value. From the three scores of wighted voting, each category class is added up and the result of the largest score will be selected as a decision. From the data testing 1, the ADHD category class that is determined is "Impulsive" because the value is 1.7234 which is greater than the number of weighted voting "Hyperactivity" which is 0.6666.

The same calculation is carried out for further testing data, namely data 2, 8, and 9. Of the 4 testing data used, an accuracy of 50% is carried out with 2 correct predictive values and 2 false predictions, namely data 2 and 9 testing. testing data is shown in table 11.

Table 11. 4 Data Testing Accuracy Calculation Results

Nilai k	Data testing	Hasil	
		Data Sebenarnya	Prediksi Sistem
3	1	Impulsif	Impulsif
	2	Hiperaktif	Impulsif
	8	Impulsif	Impulsif
	9	Hiperaktif	Impulsif

Akurasi	50%
---------	-----

#### 4. CONCLUSION

ADHD symptoms obtained from books then make a questionnaire and weighted the scores to determine a child identified with ADHD and class can use the Modified K-Nearest Neighbor algorithm and the application of the Modified K-Nearest Neighbor algorithm is a development of the modified KNN algorithm by adding validation calculations and weight voting.

#### REFERENCES

- [1] K. W. Lange, S. Reichl, K. M. Lange, L. Tucha, and O. Tucha, "The history of attention deficit hyperactivity disorder," *ADHD Atten. Deficit Hyperact. Disord.*, 2010, doi: 10.1007/s12402-010-0045-8.
- [2] M. Miller and S. P. Hinshaw, "Attention-Deficit/Hyperactivity Disorder," in *Encyclopedia of Mental Health: Second Edition*, 2016.
- [3] P. Shaw, A. Stringaris, J. Nigg, and E. Leibenluft, "Emotion dysregulation in attention deficit hyperactivity disorder," *American Journal of Psychiatry*. 2014, doi: 10.1176/appi.ajp.2013.13070966.
- [4] S. B. Sinaga, "The Application of the Certainty Factor method in Diagnosing Measles," *Login J. Teknol. Komput.*, vol. 12, no. 2, pp. 31–35, 2018, doi: <https://doi.org/10.24224/login.v12i2.24>.
- [5] S. Jiang, G. Pang, M. Wu, and L. Kuang, "An improved K-nearest-neighbor algorithm for text categorization," *Expert Syst. Appl.*, 2012, doi: 10.1016/j.eswa.2011.08.040.
- [6] G. O. Mustakim, "Algoritma K-Nearest Neighbor Classification," *J. Sains, Teknol. dan Ind.*, 2016.
- [7] H. A. Ramadhan and D. A. Putri, "Big Data, Kecerdasan Buatan, Blockchain, dan Teknologi Finansial di Indonesia," *Direktorat Jenderal Apl. Inform. Kementeri. Komun. dan Inform.*, 2018.
- [8] Muhammad Dahria, "Kecerdasan buatan ( Artificial Intelligence )," *Artif. Intell.*, 2014.
- [9] H. Leidiyana, "Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor," *J. Penelit. Ilmu Komputer, Syst. Embed. Log.*, 2013.
- [10] F. Liantoni, "Klasifikasi Daun Dengan Perbaikan Fitur Citra Menggunakan Metode K-Nearest Neighbor," *J. Ultim.*, 2016, doi: 10.31937/ti.v7i2.356.